HER2 Testing and Clinical Decision Making in Gastroesophageal Adenocarcinoma

Guideline from the College of American Pathologists, American Society for Clinical Pathology, and American Society of Clinical Oncology

Corresponding Author:
Angela N. Bartley, MD, FCAP

Authors:
Carol Colassaco, MLIS, SCT(ASCP)
Nofisat Ismaila, MD
Christina Ventura, MLS(ASCP)


*The Supplemental Digital Content was not copyedited by Archives of Pathology and Laboratory Medicine, American Journal of Clinical Pathology, or Journal of Clinical Oncology

© 2016 CAP, ASCP, ASCO. All rights reserved.
METHODS USED TO PRODUCE THE GUIDELINE

Panel Composition
The College of American Pathologists (CAP), the American Society for Clinical Pathology (ASCP), and the American Society of Clinical Oncology (ASCO) convened an Expert Panel (EP) consisting of pathologists, oncologists, gastroenterologist, and a methodologist to develop an evidence-based guideline to help establish standard HER2 testing to guide targeted therapies, and advance personalized care for patients. All three organizations appointed a representative to serve as a co-chair, with one taking a leadership role (AB). All three organizations approved the appointment of panel members. The EP and the methodologist performed the systematic evidence review. An advisory panel (AP) of pathologists, oncologists, and patient advocates also helped in the development of the guideline. The role of the AP members was to provide guidance and feedback on the key questions for the literature search, vet the draft guideline statements prior to the public comment period, and to review and provide feedback for the manuscript and supplemental digital content.

Conflict of Interest (COI) Policy
The CAP, ASCP, and ASCO require that any individual influencing the content of Practice Guidelines provide disclosure of the existence and extent of any financial interest relevant to the content of these guidelines on molecular biomarkers, tests or therapies associated with gastroesophageal adenocarcinoma. The intent of disclosures is to provide transparency regarding any relationship that may bias an individual's participation or work product of which, if known, could give the perception of bias. Disclosures of actual or perceived conflicts of interest (COI) of all members of the practice guidelines development group allow users to interpret recommendations in light of COIs. The COI policy is based on and consistent with the recommendations in the Institute of Medicine’s 2011 report, Clinical Practice Guidelines We Can Trust.1

Prior to acceptance on the expert or advisory panel, potential members completed a joint guideline conflict of interest (COI) disclosure process, whose policy and form (in effect December 2014) require disclosure of material financial interest in, or potential for benefit of significant value from the guideline’s development or its recommendations 12 months prior through the time of publication. The potential members completed the COI disclosure form, listing any relationship that could be interpreted as constituting an actual, potential, or apparent conflict. Examples of conflicts of interest with relevant commercial entities were provided to the participants using a Conflict of Interest (COI) Policy List of Affected Companies For the CAP/ASCP/ASCO HER2 Testing in Gastroesophageal Cancer document.

The CAP/ASCP/ASCO joint guideline COI policy uses the following criteria to define relationships that could be interpreted as constituting an actual, potential, or apparent conflict:
1. Stock options or bond holdings in a relevant commercial entity or self-directed pension plan
2. Research grants from a relevant commercial entity
3. Employment (full or part-time) by a relevant commercial entity
4. Ownership or partnership in relevant corporate entities, including equities and stock options
5. Consulting or advisory fees from relevant commercial entities
6. Other remuneration from relevant commercial entities, including free or discounted products or equipment, trips, accommodations, tickets to sports or entertainment events, etc.
7. Non-remunerative positions of influence in a relevant commercial entity such as officer, board member, trustee, spokesperson, advisor
8. Royalties from relevant commercial entities
9. Intellectual property rights, i.e., patents issued or pending
10. Lecture or speaker fees/honoraria from relevant commercial entities
11. Other relationships, e.g., research collaborations, to be identified with details, as needed

All project participants were required to disclose conflicts prior to beginning and continuously throughout the project’s timeline. All disclosed conflicts were reviewed by a joint COI Review Committee composed of staff officials from each of the respective organizations. The joint COI Review Committee agreed, by majority vote, on any resolution of actual or perceived conflicts of interest.

Only one of the co-chairs could receive research support from a relevant commercial entity (no other relevant relationship was allowed). At least 51% of the Expert Panel had no existing or future relationships planned with relevant commercial entities during the development and publication of the practice guidelines. For the remaining 49%, such relationships did not preclude Expert Panel membership. At the discretion of the co-chairs, these individuals were asked to recuse themselves from discussing topics and abstained from voting on any decisions or approvals relevant to their relationships. Expert panel members’ disclosed conflicts are listed in the appendix of the manuscript. Advisory panel members had a disclosure requirement, but conflicts were not subject to management by the COI Review Committee.

The CAP, ASCP and ASCO provided funding for the administration of the project; no industry funds were used in the development of the guideline. All panel members volunteered their time and were not compensated for their involvement, except for the contracted methodologist.

**Literature Review and Analysis**

The Expert Panel met face-to-face on April 25, 2015 to develop scope and key questions, and to launch the systematic review. The panel met again on August 29, 2015 to review and assess the evidence and draft the recommendations. In addition, small group of panel members met a total of 16 times via web conferences to conduct the systematic review, assess the solicited feedback from the public comment period and finalize the recommendations. Additional work was completed via electronic mail.

The expert panel formed the following key questions (KQs) on which to base the literature search:

**Clinical question 1:** What is the optimal testing algorithm for the assessment of HER2 status in patients with gastroesophageal adenocarcinoma?

1. Should HER2 testing be requested for every patient diagnosed with gastroesophageal adenocarcinoma?
2. Which of the tumor specimen(s) is the most appropriate to perform HER2 testing?
   - Biopsy specimen from primary tumor
   - Resection specimen
   - Tissue from metastatic site
   - Fine needle aspirate or cytology specimen from primary or metastatic tumor
3. In patients with HER2 positive results, under what clinical scenario should HER2 targeted therapy be initiated?
4. Should HER2 directed therapy be delayed if HER2 status cannot be confirmed as positive or negative (i.e. if an equivocal result is found with immunohistochemistry [IHC])?
5. Under what circumstances should patient samples be retested for HER2?
   - Biopsy (primary tumor) versus resection
   - Biopsy (primary tumor) versus resection versus metastatic tissue
   - Concurrent versus later metastatic tissue
6. What are the clinical performance characteristics of IHC and in situ hybridization (ISH)?

Clinical question 2: What strategies can help ensure optimal performance, interpretation and reporting of established assays in patients with gastroesophageal adenocarcinoma?

7. What are the analytic performance characteristics of IHC and ISH (e.g. sensitivity, specificity, reproducibility, gold standard, consensus among testing laboratories)?
   - What is the correlation between different IHC scores (0-3) and ISH results?
8. What are the acceptable methodologies for HER2 IHC (different antibodies) and ISH (different probes platforms)?
9. What is the optimal testing algorithm for the assessment of HER2 status?
   - Which testing modality or algorithm is most cost effective?
   - When and how should reflex (ISH) testing be done?
10. What are the steps/procedures needed to analytically validate a laboratory developed HER2 gastroesophageal adenocarcinoma assay before reporting results on patient samples?
   - Should different validation be performed in gastroesophageal adenocarcinoma and breast specimen?
11. What is the best scoring method for IHC and ISH in gastroesophageal adenocarcinoma specimens?
   - Can HER2 copy numbers be used to define HER2 status in addition to HER2 and chromosome enumeration probe 17 (CEP17) ratios (i.e. in cases with apparent polysomy) in ISH testing as a positive result?
   - Should the scoring criteria be the same for biopsy specimen versus resection specimen?
   - How should HER2 heterogeneity be interpreted and/or reported?
   - When should a specimen be reported as indeterminate?
12. How should HER2 results be reported?
13. What is adequate specimen handling for gastroesophageal adenocarcinoma testing?
14. What is the appropriate morphologic correlation for interpretation of ISH?
15. What are the optimal quality assurance/quality control standards that labs should adhere to? (e.g. proficiency testing, laboratory volume, ongoing personnel training, appropriate quality control)
16. Is there a role for HER2 genomic testing?

All expert panelists participated in the systematic evidence review. The title-abstract review was primarily reviewed by the methodologist with the assistance of the co-chairs. The full text review was performed in duplicate by two members of the expert panel. The data was extracted by the methodologist and audited by members of the expert panel. All expert panelists and the methodologist performed adjudication of the conflicts. Articles meeting the inclusion criteria were assessed for strength of evidence, methodological rigor, and confirmation of validity by the methodologist. Supplemental Figure 1 displays the results of the literature review. All articles were available as discussion or background references. All members of the expert panel participated in developing draft recommendations, the assignment of the strength of recommendations based on the extracted evidence, reviewing open comment feedback, finalizing and approving final recommendations and writing/editing of the manuscript.
Peer Review

A public open comment period was held from December 8, 2015 through January 11, 2016. Twenty draft statements were posted online on the ASCP Web site www.ascp.org. The open comment period was publicized via joint society communications announcements and the following societies were deemed to have interest:

- College of American Pathologists (CAP)
- American Society for Clinical Pathology (ASCP)
- American Society for Clinical Oncology (ASCO)
- Association for Molecular Pathology (AMP)
- Association of Directors of Anatomic and Surgical Pathology (ADASP)
- Arthur Purdy Stout Society (APSS)
- Association of Pathology Chairs (APC)
- Canadian Association of Pathologists (CAP-APC)
- United States & Canadian Academy of Pathology (USCAP)
- Quality Initiative in Interpretive Pathology (QIIP) Canadian Partnership Against Cancer
- Society to Improve Diagnoses in Medicine (SIDM)
- Roger G. Haggitt Gastrointestinal Pathology Society (GIPS)
- European Society for Medical Oncology (ESMO)
- American Association for Clinical Chemistry (AACC)
- American College of Medical Genetics and Genomics (ACMG)
- Association of Community Cancer Centers (ACCC)
- National Comprehensive Cancer Network (NCCN)
- American Cancer Society
- Partnership Against Cancer American Cancer Society
- Cancer Research and Prevention Foundation
- Cancer Leadership Council
- Union for International Cancer Control
- Fight Colorectal Cancer
- Colon Cancer Alliance
- US Food and Drug Administration (FDA)
- Centers for Medicare & Medicaid Services (CMS)
- Centers for Disease Control and Prevention (CDC)
- Veteran’s Affairs (VA) and Department of Defense (DOD)

The website received 294 comments in total (Agree as written, Agree with suggested modification and Disagree responses were captured). All draft recommendations achieved between 82% to 95% agreement as written. Teams of 2 of expert panel members were assigned the draft statements for 2-3 key questions. The teams reviewed all comments received and provide an overall summary to the rest of the panel. Following panel discussion, and the final quality of evidence assessment, the panel members determined whether to maintain the original draft recommendation as is, revise it with minor language change, or consider it as a major recommendation change. The panel modified 1 draft recommendation and combined 4 draft recommendations based on the feedback from the public comment period and the panel’s discussion and considered judgment process. The panel decided that general recommendations about quality assurance, turnaround time, and specimen handling were best suited as part of the discussion, and would be included in the body of the manuscript rather than as formal recommendations. Resolution of all changes was obtained by majority consensus of the panel using nominal group technique (rounds of email discussion and multiple edited recommendations) amongst the panel members. The final recommendations were approved by the expert panel with a formal vote. The panel considered the risks and benefits throughout the whole process in their considered judgment process. Formal cost analysis or cost effectiveness was not performed.
Organizational review was instituted to review and approve the guideline. An independent review panel (IRP) representing the CAP Council on Scientific Affairs was nominated to review and approve the guideline. The CAP IRP was masked to the expert panel and vetted through a COI process. ASCP assigned the review to a Special Review Panel at the discretion of the ASCP Executive Office and the Board of Directors. The ASCO approval process required the review and approval by the Clinical Practice Guidelines Committee.

**Dissemination Plans**
Final dissemination of the guideline will be a joint process between the three organizations. There are plans to host a resource page which will include a link to the manuscript and supplement, summary of the recommendations, social media as well as patient information guides. The guideline will be promoted and presented at various society meetings.

**Systematic Evidence Review (SER)**
The objective of the SER was to develop an evidence-based guideline to determine what the optimal testing algorithm is for the assessment of HER2 status, and to determine strategies that can help ensure optimal performance, interpretation and reporting of established assays in patients with gastroesophageal adenocarcinoma. The guideline was developed to help establish standards for HER2 testing in gastroesophageal adenocarcinoma to help guide targeted therapies, and advance personalized care for patients. The scope of the SER and the KQs were established by the EP in consultation with the methodologist prior to beginning the literature search.

**Search and Selection**
A comprehensive search for literature was performed in MEDLINE using the OvidSP (5/29/2015) and PubMed (6/4/2015) interfaces. The initial MEDLINE search encompassed the publication dates of 1/1/2008 to 5/29/2015 (OvidSP) and 1/1/2008 to 6/4/2015 (PubMed). A supplemental literature search was performed utilizing Scopus (6/4/2015 to identify relevant articles published in journals not indexed in MEDLINE and published between 1/1/2008 and 6/4/2015. The literature search of the electronic databases was conducted in two arms – one combined medical subject headings (MeSH) and keywords to address the concepts of esophagogastric neoplasms, Her-2/ErbBB-2, and therapy (e.g., monoclonal antibodies/antineoplastic agents/molecular targeted therapy), and the second combined MeSH terms and keywords for esophagogastric neoplasms, Her-2/ErbBB-2 and laboratory testing methods. The results of both arms of the search were combined and deduplicated.

In addition to the searches of electronic databases, a search for grey (unindexed) literature was completed that included a review of guideline repository sites (e.g., Agency for Healthcare Research and Quality, Guidelines International Network), the Cochrane Library, Prospero, and relevant organizations’ websites.

The Ovid, PubMed, and Scopus search strategies are included as Supplemental Figure 2.

Selection at all levels was based on predetermined inclusion/exclusion criteria.

**Included were:**
1. Patients with gastroesophageal adenocarcinoma
2. Patients of all ages
3. Male and female patients
4. Patients with any stage of disease and tumors of any grade
5. Human studies
6. Studies published in English
7. Studies that met the defined study design requirements
8. Studies that addressed at least one of the key questions

**Excluded were:**
1. Patients with all other tumor primaries and types are excluded, including esophageal squamous cell carcinomas
2. Patients with noninvasive tumors (intraepithelial, dysplasia, in situ, polyps without carcinoma) are excluded
3. Non-English language articles
4. Animal studies
5. Studies published prior to 2008
6. Studies that did not meet the defined study design requirements
7. Studies that did not address at least one of the defined inclusion criteria

Outcomes of Interest
The primary outcomes of interest included survival outcomes and performance characteristics of laboratory testing assays. Survival outcomes included: overall survival (OS), disease-free survival (DFS), progression free survival (PFS), response rate, recurrence-free survival, time to recurrence, response to therapy (e.g., complete and partial response). Laboratory data and test performing characteristics included sensitivity and specificity of testing methods, and concordance.

Data Extraction & Management
Following the initial search, titles and abstracts of retrieved studies were reviewed by the methodologist and co-chairs for relevancy. Conflicts were resolved by initial reviewers and further adjudicated by a project co-chair, if necessary. Titles and abstracts advanced to full text review if the screener felt the study was relevant to the guideline, the laboratory was laboratory-focused or clinically-focused based on the population of interest and the intervention or test of interest, and the article met the established study design specifications:

For Clinical studies:
   Included were:
      1. Systematic reviews with or without meta-analyses
      2. Other reviews (consensus, expert panel, guidelines)
      3. Randomized trials (Phase II or III, placebo-controlled, blinded)

   Excluded were:
      1. Phase I randomized trials
      2. Non-randomized controlled trials
      3. Uncontrolled trials
      4. Observational studies
      5. Non-comparative studies (case reports, case series, time series)
      6. Follow-up studies
      7. Qualitative studies
      8. Mixed methods studies
      9. Narrative reviews
      10. Meeting abstracts

For Laboratory studies:
   Included were:
      1. Systematic reviews with or without meta-analyses
      2. Other reviews (consensus, expert panel, guidelines)
      3. Randomized trials (Phase I, II, III, placebo-controlled, blinded)
      4. Non-randomized controlled trials
      5. Uncontrolled trials
      6. Observational studies

   Excluded were:
      1. Follow-up studies
      2. Qualitative studies
3. Mixed methods studies
4. Time series – non-comparative studies
5. Meeting abstracts
6. Books, letters, editorials

Full text articles were reviewed for relevancy by two expert panel members to determine eligibility, and conflicts were resolved by the initial reviewers and further adjudicated by a project co-chair, if necessary. In cases of duplication of reporting study results, the most inclusive were retained. Articles advanced to data extraction if they addressed at least one of the key questions, contained measurable data, and were within the project’s scope and met the previously described inclusion/exclusion criteria. Data extraction was performed by a methodologist and audited by one expert panel member. Any discrepancies in data extraction were resolved by discussion. A bibliographic database was established in DistillerSR (Ontario, Canada) and EndNote (Thomson Reuters, Carlsbad, CA) to track all literature identified and reviewed during the study.

Quality Assessment Methods
An assessment of the quality of the evidence was performed for all retained studies following application of the inclusion and exclusion criteria. Using this method, studies deemed be of low quality would not be excluded from the systematic review, but would be retained and their methodological strengths and weaknesses discussed where relevant. Studies would be assessed by confirming the presence of items related to both internal and external validity, and which are all associated with methodological rigor and a decrease in the risk of bias. These items were assessed as being either yes, no, partial, not reported (NR), or not applicable (N/A) in the following way:

Systematic Reviews (SRs) and Meta-analyses were assessed for quality by confirming the following attributes were considered and incorporated in its design as recommended by the Institute of Medicine (IOM).1 (Summarized in Supplemental Table 1)

- Included a multidisciplinary panel
- Patient preferences were considered
- Important patient sub-types were considered
- Methods were well-described and reproducible
- Information on potential conflicts of interest were gathered and disclosed
- Quality of the evidence was assessed
- Strength of the evidence was rated
- Sources of funding are disclosed

Meta-analyses (M-As) were assessed in a similar fashion to SRs:

- Based on a systematic review
- Methods were well-described and reproducible
- Quality of the evidence was assessed
- Any planned pooling was stated a priori
- Limitations of the analysis are discussed
- Sources of funding are disclosed

Randomized Control Trials (RCTs) and Quasi-RCTs were assessed for quality according to reporting and full description of:

- Randomization method fully-described
- Treatment allocation was concealed
- Sample size was sufficient
- Validated and reliable measures
- Details on any blinding was provided
• Provided details of all planned analyses
• Stated the expected effect size and described the statistical power calculation
• Reported the length of follow-up
• Provided a description of the baseline characteristics for all patients by treatment/assessment arm
• Sources of funding are disclosed

Non-randomized clinical trials (NRCTs), prospective cohort studies (PCS), and retrospective cohort studies (RCS) were assessed according to:

• Balance between treatment/assessment groups
• Reporting of baseline characteristics
• Reporting if any adjustments were made where baseline differences were detected
• Sources of funding

Supplemental Table 1-6 summarizes the quality assessment results by study design and overall risk of bias assessment.

**Strength of Recommendations**
The expert panel reviewed all the synthesized evidence and drafted recommendations during one of the in-person meetings. For each recommendation, there was a discussion on the quality of the evidence available, the harms versus benefits, values, as well as limitations. The strength of recommendations designations and rationale are listed in Supplemental Table 7.
## Quality Assessment Results by Study Design

### Supplemental Table 1: Systematic reviews (N=1)

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Multi-disciplinary panel</th>
<th>Patient preferences considered</th>
<th>Important patient sub-types considered</th>
<th>Well-described and reproducible methods</th>
<th>COI's are examined</th>
<th>Rated quality of the Evidence</th>
<th>Rated strength of the evidence</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chua, 2012</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>No</td>
<td>Unsure/insufficient detail/NR</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

Abbreviations: COI, conflict of interest; NR, not reported

### Supplemental Table 2 Meta-analyses (N=2)

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Based on systematic review</th>
<th>Reproducible methods</th>
<th>Quality assessment of included studies</th>
<th>Planned pooling stated a priori</th>
<th>Limitations of the study</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peng, 2015</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Wang, 2011</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsue/insufficient detail/NR</td>
<td>No</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

Abbreviations: NR, not reported

### Supplemental Table 3: Randomized control trial (N=2)

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Adequate Randomization</th>
<th>Concealed allocation</th>
<th>Sufficient Sample Size</th>
<th>Similar groups</th>
<th>Blinded Validated and Reliable measures</th>
<th>Adequate follow up</th>
<th>ITT</th>
<th>Insignificant COIs</th>
<th>Overall potential Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bang, 2010</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Van Cutsem, 2015</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Abbreviations: COI, conflict of interest; ITT, intention to treat; NR, not reported.
### Supplemental Table 4: Prospective cohort (N=27)

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinugas a, 2015&lt;sup&gt;7&lt;/sup&gt;</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Ge, 2015&lt;sup&gt;8&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Wang, 2015&lt;sup&gt;9&lt;/sup&gt;</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Gumusa y, 2015&lt;sup&gt;10&lt;/sup&gt;</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Qiu, 2015&lt;sup&gt;11&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Cappelle sso, 2015&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Selckbi rick, 2014&lt;sup&gt;13&lt;/sup&gt;</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Wong, 2015&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Stanek, 2014&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Huang*, 2013&lt;sup&gt;16&lt;/sup&gt;</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Kushima , 2014&lt;sup&gt;17&lt;/sup&gt;</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Wang, 2014&lt;sup&gt;18&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
</tbody>
</table>

* Epidemiological study. Abbreviation: NR, not reported
<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tajiri, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Wang, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Kochi, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Ormenisan, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Selcukbicirik, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Low</td>
</tr>
<tr>
<td>He, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Janjigian, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Tamura, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>High</td>
</tr>
<tr>
<td>Halon, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Cho, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Fox, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Liu, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Schoppmann, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Choritz, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Sekaran, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
</tbody>
</table>

Abbreviation: NR, not reported
### Supplemental Table 6: Retrospective cohort (N=15)

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schmitt, 2015</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Stahl, 2015</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Van Hagen, 2015</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Ieni, 2014</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Rakhshani, 2014</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Unsure/insufficient detail/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Kumara Singhge, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>High</td>
</tr>
<tr>
<td>Geng, 2014</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Kimura, 2014</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Cho, 2013</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Pala, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Lee, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Grabsch, 2010</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Song, 2010</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

Abbreviation: NR, not reported
### Supplemental Table 7: Retrospective cohort (N=15), continued

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Park, 2015</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Gullo, 2015</td>
<td>No</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

Abbreviation: NR, not reported.

### Supplemental Table 8: Prospective-Retrospective studies (N=69)

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otsu, 2015</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Feuchtinger, 2015</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Kim, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Chen, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Grillo, 2013</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Tafe, 2015</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>High</td>
</tr>
<tr>
<td>Koopman, 2015</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Sheffield, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Chen, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Werner, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Yoshida, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable; NR, not reported.
### Supplemental Table 9: Prospective-Retrospective studies (N=69), continued

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geppert, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Aizawa, 2014</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Prins, 2014</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Gomez-Martin, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Grin, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Abrahao-Machado, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Gasljevic, 2013</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Gordon, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Han, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Fusco, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Cruz-Reyes, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Shan, 2013</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Cho, 2013</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Prins, 2013</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Okines, 2013</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable; NR, not reported
### Supplemental Table 10: Prospective-Retrospective studies (N=69), continued

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phillips, 2013(^75)</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Pirrelli, 2013(^76)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Warneke, 2013(^77)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Pagni, 2013(^78)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Zhou, 2012(^79)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Hirschman, 2012(^80)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Yoon, 2012(^81)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Terashima, 2012(^82)</td>
<td>Yes</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Kiyose, 2012(^83)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Fassan, 2012(^84)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Asioli, 2012(^85)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Gomez-Martin, 2012(^86)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Fassan, 2012(^87)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Jeung, 2012(^88)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Kim, 2012(^89)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Radu, 2012(^90)</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable; NR, not reported
<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sornmayura, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Park, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Mrklic, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Tsapralis, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Yoon, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Yang, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Kunz, 2012</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Hsu, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Lee, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Kim, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Tafe, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Low</td>
</tr>
<tr>
<td>Yan, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>High</td>
</tr>
<tr>
<td>Kim, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Garcia-Garcia, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Langer, 2011</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable; NR, not reported
<table>
<thead>
<tr>
<th>Author, year</th>
<th>Was there balance between treatment/assessment groups?</th>
<th>Reporting of baseline characteristics (and any differences detected between groups)</th>
<th>Reporting of any adjustment when differences were present</th>
<th>Funding source</th>
<th>Overall risk of bias assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hu, 2011106</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Thompson, 2011107</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Im, 2011108</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Bozzetti, 2011109</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Yan, 2011110</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Moelans, 2011111</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Boers, 2011112</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Schoppman, 2010113</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Yan, 2010114</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Marx, 2009115</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/NA/NR</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Barros-Silva, 2009116</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Unsure/insufficient detail/NR</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Xie, 2009117</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>Unsure/insufficient detail/NR</td>
<td>No</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable; NR, not reported
### Supplemental Table 13: Strength of Recommendations

<table>
<thead>
<tr>
<th>CAP Designation</th>
<th>GLIDES Designation</th>
<th>Recommendation</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong Recommendation</td>
<td>Strong</td>
<td>Recommend For or Against a particular practice (Can include must or should)</td>
<td>Supported by high (convincing) or intermediate (adequate) quality of evidence and clear benefit that outweighs any harms</td>
</tr>
<tr>
<td>Recommendation</td>
<td>Moderate</td>
<td>Recommend For or Against a particular practice (Can include should or may)</td>
<td>Some limitations in quality of evidence (intermediate [adequate] or low [inadequate]), balance of benefits and harms, values, or costs but panel concludes that there is sufficient evidence and/or benefit to inform a recommendation.</td>
</tr>
<tr>
<td>Expert Consensus Opinion</td>
<td>Weak</td>
<td>Recommend For or Against a particular practice (Can include should or may)</td>
<td>Serious limitations in quality of evidence (low [inadequate] or insufficient), balance of benefits and harms, values or costs, but panel consensus is that a statement is necessary.</td>
</tr>
<tr>
<td>No Recommendation</td>
<td>N/A</td>
<td>No recommendation for or against a particular practice</td>
<td>Insufficient evidence or agreement of the balance of benefits and harms, values, or costs to provide a recommendation</td>
</tr>
</tbody>
</table>

Data derived from Guyatt et al.\(^{118}\) Abbreviations: CAP, College of American Pathologists; GLIDES, Guidelines into Decision Support (Yale University, New Haven, Connecticut); N/A, not applicable
Supplemental Figure 1. Literature Review Flow Diagram


*Additional searches from Cochrane, NICE, Prospero, expert panel input

**Records excluded at title-abstract screening, with reasons (N = 689): Not the intervention of test of interest (N=288); Not the population of interest (N=114); Reviews, case reports, letters, editorials, books (N=218); Esophageal squamous cell carcinoma (N=24); Consensus document, opinion papers (N=16); For Clinical studies: Phase 1 RCT and observational (N=29)
Supplemental Figure 2: Literature search strategies

Ovid Search Strategy

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1946 to Present>, Ovid MEDLINE(R) Daily Update <May 29, 2015>

--------------------------------------------------------------------------------
1     stomach neoplasms/ (77111)
2     esophageal neoplasms/ (40064)
3     esophagogastric junction/ (6371)
4     stomach/ (50906)
5     Barrett esophagus/ (6496)
6     adenocarcinoma/ (130389)
7     carcinoma/ (66303)
8     (stomach or gastric or esophagogastric or gastro?esophageal or gastro?oesophageal or Barrett$ or oesophag$ or esophag$).tw. (369432)
9     (adeno$ or cancer or carcinoma$ or neoplasm$ or malignan$ or tumo?r$).tw. (2524215)
10    6 or 7 or 9 (2541448)
11    3 or 4 or 5 or 8 (384817)
12    10 and 11 (132621)
13    1 or 2 or 12 (157632)
14    Genes, erbB-2/ (2755)
15    Receptor, ErbB-2/ (17248)
16    (HER?2$ or ERBB?2).tw. (18041)
17    "human epidermal growth factor receptor 2".tw. (3230)
18    or/14-17 (27181)
19    Antibodies, Monoclonal, Humanized/ (27281)
20    antibodies, monoclonal/ (170319)
21    exp antineoplastic agents/ (851636)
22    protein kinase inhibitors/ (27414)
23    quinazolines/ (14709)
24    quinolines/ (18731)
25    maytansine/ (372)
26    pertuzumab.nm. (254)
27    lapatinib.nm. (1042)
28    BIBW 2992.nm. (154)
29    trastuzumab.nm. (4419)
30    ado-trastuzumab emtansine.nm. (123)
31    molecular targeted therapy/ (11581)
32    ((molecular or targeted or directed) and (treat$ or therap$ or protocol)).tw. (330939)
33    (trastuzumab or lapatinib or pertuzumab or ado?trastuzumab or T?DM1 or TDM?1 or afatinib or neratinib or Tykerb or Herceptin or Kadcyla or Perjeta or Giotrif or HKI?272 or Hercon or Giotrif or Tomtovok or Tovok).tw. (7523)
34    or/19-33 (1321620)
35    immunohistochemistry/ (255112)
36    fluorescent antibody technique/ (92954)
37    fluorescent antibody technique, direct/ (2784)
38    fluorescent antibody technique, indirect/ (15517)
39    in situ hybridization, fluorescence/ (36313)
40    exp genetic techniques/ (1575944)
41    (FISH or ISH or CISH or SISH or DISH or hybrid#ation or fluorescent$ or probe$ or platform$ or algorithm or modalit$).tw. (1056211)
PubMed Search Strategy

(((immunohistochemistry[mh:noexp]) OR "fluorescent antibody technique", indirect[mh:noexp]) OR "fluorescent antibody technique, direct"
[mh:noexp]) OR "in situ hybridization, fluorescence"[mh:noexp]) OR "genetic
techniques"[MeSH Terms] OR FISH[Title/Abstract]) OR ISH[Title/Abstract]) OR CISH[Title/Abstract]) OR SISH[Title/Abstract]) OR DISH[Title/Abstract]) OR hybridization[Title/Abstract]) OR
hybridisation[Title/Abstract]) OR fluorescence[Title/Abstract]) OR probe[Title/Abstract]) OR
probes[Title/Abstract]) OR platform[Title/Abstract]) OR platforms[Title/Abstract]) OR
algorithm[Title/Abstract]) OR modality[Title/Abstract]) OR immunohistochemistry[Title/Abstract]) OR
immunocytochemistry[Title/Abstract]) OR immunohistochemical[Title/Abstract]) OR
immunocytotoxic[Title/Abstract]) OR IHC[Title/Abstract]) OR ICC[Title/Abstract]) OR
assay[Title/Abstract]) OR assays[Title/Abstract]) OR antibody[Title/Abstract]) OR
antibodies[Title/Abstract]) OR sequencing[Title/Abstract]) OR sequence[Title/Abstract]) OR
"copy number"[Title/Abstract]) OR polymorphism[Title/Abstract]) OR
immunoprecipitation[Title/Abstract]) OR "image analysis"[Title/Abstract]) OR "image analyse"
[Title/Abstract]) AND (((trastuzumab[Title/Abstract]) OR lapatinib[Title/Abstract]) OR
dutchtrastuzumab[Title/Abstract]) OR adotrastuzumab[Title/Abstract] OR "ado
trastuzumab"[Title/Abstract]) OR T-DM1[Title/Abstract]) OR TDM1[Title/Abstract]) OR
"treatment"[Title/Abstract]) OR therapy[Title/Abstract]) OR therapies[Title/Abstract]) OR
protocol[Title/Abstract]) OR protocols[Title/Abstract]) AND
"molecular targeted therapy"
[Title/Abstract]) OR targeted[Title/Abstract]) OR directed[Title/Abstract])) OR
"antibodies, monoclonal, humanized"[mh:noexp]) OR "antibodies, monoclonal"
[mh:noexp]) OR "antineoplastic agents"[MeSH Terms] OR "protein kinase inhibitors"
[mh:noexp]) OR
quinazolines[mh:noexp]) OR quinolines[mh:noexp]) OR maytansine[mh:noexp]) OR
pertuzumab[Supplementary Concept]) OR lapatinib[Supplementary Concept]) OR BIWB
2992[Supplementary Concept]) OR trastuzumab[Supplementary Concept]) OR ado
trastuzumab emtansine[Supplementary Concept]) OR "molecular targeted therapy"
["mh:noexp") AND
"adenocarcinoma[MeSH Terms]) OR carcinoma[MeSH Terms]) OR cancers[Title/Abstract]) OR
cancer[Title/Abstract]) OR carcinomas[Title/Abstract]) OR carcinoma[Title/Abstract]) OR
adenocarcinomas[Title/Abstract]) OR neoplasm[Title/Abstract]) OR
"malignancy[Title/Abstract]) OR malignancy[Title/Abstract]) OR
malignancies[Title/Abstract]) OR tumor[Title/Abstract]) OR tumors[Title/Abstract]) OR
tumours[Title/Abstract]) OR tumour[Title/Abstract]) OR stomach[Title/Abstract]) OR
gastroesophageal[Title/Abstract]) OR oesophageal[Title/Abstract]) OR
oesophageal[Title/Abstract]) OR stomach[Title/Abstract]) OR oesopagus[Title/Abstract]) OR
stomach[MeSH Terms]) OR
esophagogastric junction[MeSH Terms]) OR Barrett esophagus[MeSH Terms])))) AND (((((Genes, erbB-2[MeSH Terms]) OR "Receptor, ErbB 2"[MeSH Terms]) OR (HER2[Title/Abstract] OR HER-2[Title/Abstract] OR ERBB2[Title/Abstract] OR ERBB-2[Title/Abstract])) OR "human epidermal growth factor receptor 2"[Title/Abstract]))) AND ("2008/01/01"[PDat] : "2015/12/31"[PDat] ) AND Humans[Mesh]) NOT ((comment[Publication Type] OR editorial[Publication Type] OR letter[Publication Type])) NOT ((comment[Publication Type] OR editorial[Publication Type] OR letter[Publication Type]))
REFERENCES


